

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: 2003年 3月12日

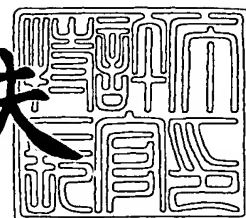
出願番号
Application Number: 特願2003-066933
[ST. 10/C]: [JP2003-066933]

出願人
Applicant(s): 株式会社エヌ・ティ・ティ・ドコモ
古井 貞熙

2004年 2月 9日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



出証番号 出証特2004-3007931

【書類名】 特許願

【整理番号】 003010

【提出日】 平成15年 3月12日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 15/00

【発明者】

【住所又は居所】 東京都千代田区永田町二丁目 1 1 番 1 号 株式会社エヌ
・ ティ ・ ティ ・ ドコモ内

【氏名】 張 志鵬

【発明者】

【住所又は居所】 東京都千代田区永田町二丁目 1 1 番 1 号 株式会社エヌ
・ ティ ・ ティ ・ ドコモ内

【氏名】 大辻 清太

【発明者】

【住所又は居所】 東京都千代田区永田町二丁目 1 1 番 1 号 株式会社エヌ
・ ティ ・ ティ ・ ドコモ内

【氏名】 杉村 利明

【発明者】

【住所又は居所】 東京都世田谷区新町二丁目 3 6 番 6 号

【氏名】 古井 貞熙

【特許出願人】

【識別番号】 392026693

【氏名又は名称】 株式会社エヌ・ティ・ティ・ドコモ

【特許出願人】

【識別番号】 502347788

【氏名又は名称】 古井 貞熙



【代理人】

【識別番号】 100066980

【弁理士】

【氏名又は名称】 森 哲也

【選任した代理人】

【識別番号】 100075579

【弁理士】

【氏名又は名称】 内藤 嘉昭

【選任した代理人】

【識別番号】 100103850

【弁理士】

【氏名又は名称】 崔 秀▲てつ▼

【手数料の表示】

【予納台帳番号】 001638

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0016816

【プルーフの要否】 要



【書類名】 明細書

【発明の名称】 音声モデルの雑音適応化システム、雑音適応化方法、及び、音声認識雑音適応化プログラム

【特許請求の範囲】

【請求項 1】 クリーンな音声データを用いて学習した不特定雑音用の音声モデルを、認識対象となる雑音環境の音声に適応化する音声モデルの雑音適応化システムであって、雑音が重畳された雑音重畳音声のクラスタリングを行うクラスタリング手段と、このクラスタリング手段のクラスタリング結果に基づいて木構造の雑音音声モデル空間を作成する音声モデル空間作成手段と、音声認識対象となる入力雑音音声の音声特徴パラメータを抽出するパラメータ抽出手段と、前記音声モデル空間作成手段によって作成された木構造の雑音音声モデル空間から最適なモデルを選択する選択手段と、この選択手段によって選択されたモデルについて尤度が更に大きくなるように線形変換を行う線形変換手段とを含むことを特徴とする音声モデルの雑音適応化システム。

【請求項 2】 前記クラスタリング手段は、雑音対信号比条件に合わせて前記雑音を音声に重畳することにより前記雑音重畳音声を作成し、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い、各雑音重畳音声のガウス分布モデルを作成し、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより、クラスタリング結果を得ることを特徴とする請求項 1 記載の音声モデルの雑音適応化システム。

【請求項 3】 前記選択手段は、前記パラメータ抽出手段によって抽出された音声特徴パラメータに対する尤度が最大となるモデルを選択することを特徴とする請求項 1 又は 2 記載の雑音適応化システム。

【請求項 4】 前記選択手段は、前記木構造の雑音音声モデル空間について、上層から下層に向かってたどることによってモデルを選択することを特徴とする請求項 3 記載の雑音適応化システム。

【請求項 5】 前記線形変換手段は、前記選択手段によって選択されたモデルに基づいて尤度が大きくなるように線形変換を行うことを特徴とする請求項 1 乃至 4 のいずれか 1 項に記載の雑音適応化システム。



【請求項 6】 クリーンな音声データを用いて学習した不特定雑音用の音声モデルを、認識対象となる雑音環境の音声に適応化する音声モデルの雑音適応化方法であって、雑音が重畳された雑音重畳音声のクラスタリングを行うクラスタリングステップと、このクラスタリングステップによるクラスタリング結果に基づいて木構造の雑音音声モデル空間を作成する音声モデル空間作成ステップと、音声認識対象となる入力雑音音声の音声特徴パラメータを抽出するパラメータ抽出ステップと、前記音声モデル空間作成ステップにおいて作成された木構造の雑音音声モデル空間から最適なモデルを選択する選択ステップと、この選択ステップにおいて選択されたモデルについて尤度が更に大きくなるように線形変換を行う線形変換ステップとを含むことを特徴とする音声モデルの雑音適応化方法。

【請求項 7】 前記クラスタリングステップにおいては、雑音対信号比条件に合わせて前記雑音を音声に重畳することにより前記雑音重畳音声を作成し、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い、各雑音重畳音声のガウス分布モデルを作成し、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより、クラスタリング結果を得ることを特徴とする請求項 6 記載の音声モデルの雑音適応化方法。

【請求項 8】 前記選択ステップにおいては、前記パラメータ抽出ステップにおいて抽出された音声特徴パラメータに対する尤度が最大となるモデルを選択することを特徴とする請求項 6 又は 7 記載の音声モデルの雑音適応化方法。

【請求項 9】 前記選択ステップにおいては、前記木構造の雑音音声モデル空間について、上層から下層に向かってたどることによってモデルを選択することを特徴とする請求項 8 記載の音声モデルの雑音適応化方法。

【請求項 10】 前記線形変換ステップにおいては、前記選択ステップにおいて選択されたモデルに基づいて尤度が大きくなるように線形変換を行うことを特徴とする請求項 6 乃至 9 のいずれか 1 項に記載の音声モデルの雑音適応化方法。

【請求項 11】 コンピュータを制御することにより、クリーンな音声データを用いて学習した不特定雑音用の音声モデルを、認識対象となる雑音環境の音声に適応化するための音声認識雑音適応化プログラムであって、雑音が重畳され



た雑音重畳音声のクラスタリングを行うクラスタリングステップと、このクラスタリングステップによるクラスタリング結果に基づいて木構造の雑音音声モデル空間を作成する音声モデル空間作成ステップと、音声認識対象となる入力雑音音声の音声特徴パラメータを抽出するパラメータ抽出ステップと、前記音声モデル空間作成ステップにおいて作成された木構造の雑音音声モデル空間から最適なモデルを選択する選択ステップと、この選択ステップにおいて選択されたモデルについて尤度が更に大きくなるように線形変換を行う線形変換ステップとを含むことを特徴とする音声認識雑音適応化プログラム。

【請求項 12】 前記クラスタリングステップにおいては、雑音対信号比条件に合わせて前記雑音を音声に重畳することにより前記雑音重畳音声を作成し、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い、各雑音重畳音声のガウス分布モデルを作成し、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより、クラスタリング結果を得ることを特徴とする請求項 11 記載の音声認識雑音適応化プログラム。

【請求項 13】 前記選択ステップにおいては、前記パラメータ抽出ステップにおいて抽出された音声特徴パラメータに対する尤度が最大となるモデルを選択することを特徴とする請求項 11 又は 12 記載の音声認識雑音適応化プログラム。

【請求項 14】 前記選択ステップにおいては、前記木構造の雑音音声モデル空間について、上層から下層に向かってたどることによってモデルを選択することを特徴とする請求項 13 記載の音声認識雑音適応化プログラム。

【請求項 15】 前記線形変換ステップにおいては、前記選択ステップにおいて選択されたモデルに基づいて尤度が大きくなるように線形変換を行うことを特徴とする請求項 11 乃至 14 のいずれか 1 項に記載の音声認識雑音適応化プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は音声モデルの雑音適応化システム、雑音適応化方法、及び、音声認識

雑音適応化プログラムに関し、特に音声の特徴を隠れマルコフモデル (Hidden Markov Model; 以下HMMと略称する) によってモデル化したクリーンな音声モデルを、認識対象となる雑音音声を用いて、その雑音環境に対する認識率を高めるように適応化する音声モデルの雑音適応化システム、雑音適応化方法、及び、音声認識雑音適応化プログラムに関する。

【0002】

【従来の技術】

従来の木構造区分線形変換手法として、非特許文献1がある。非特許文献1に記載されている手法に従うと、雑音のクラスタリングを行い、このクラスタリング結果に基づいて木構造の雑音音声モデル空間を作成し、音声認識対象となる入力雑音音声の音声特徴パラメータを抽出し、木構造の雑音音声モデル空間から最適なモデルを選択し、選択されたモデルについて尤度が更に大きくなるように線形変換を行うことにより、入力音声に対する音声モデル系列の推定をより正確に行うことができる。

【0003】

【非特許文献1】

張 志鵬 他著「区分線形変換による雑音適応化法における木構造クラスタリングの効果」日本音響学会2002年秋季発表会、pp. 29-30

【0004】

【発明が解決しようとする課題】

上述した非特許文献1による雑音重畳音声モデルの作成過程においては、雑音データのみに対しクラスタリングを行い、次に雑音を音声に重畳し、雑音重畳音声モデルを学習することになる。このように処理を行う場合、クラスタリング過程の雑音における特性とモデル学習過程における雑音重畳音声モデル特性とに不一致が生じるという問題がある。

【0005】

本発明は上述した従来技術の欠点を解決するためになされたものであり、その第1の目的は多くの雑音データについて最適にクラスタリング処理をすることのできる音声モデルの雑音適応化システム、雑音適応化方法、及び、音声認識雑音

適応化プログラムを提供することである。

また、本発明の第2の目的は、上記クラスタリング結果を利用することにより、音声認識率を高めることのできる音声モデルの雑音適応化システム、雑音適応化方法、及び、音声認識雑音適応化プログラムを提供することである。

【0006】

【課題を解決するための手段】

本発明の請求項1による音声モデルの雑音適応化システムは、クリーンな音声データを用いて学習した不特定雑音用の音声モデルを、認識対象となる雑音環境の音声に適応化する音声モデルの雑音適応化システムであって、雑音が重畳された雑音重畳音声のクラスタリングを行うクラスタリング手段と、このクラスタリング手段のクラスタリング結果に基づいて木構造の雑音音声モデル空間を作成する音声モデル空間作成手段と、音声認識対象となる入力雑音音声の音声特徴パラメータを抽出するパラメータ抽出手段と、前記音声モデル空間作成手段によって作成された木構造の雑音音声モデル空間から最適なモデルを選択する選択手段と、この選択手段によって選択されたモデルについて尤度が更に大きくなるように線形変換を行う線形変換手段とを含むことを特徴とする。このように、クラスタリング過程とモデル学習過程とを、一貫して雑音重畳音声に対して行うことによって、多くの雑音データについて最適にクラスタリング処理をすることができると共に、入力音声に対する音声モデル系列の推定をより正確に行うことができる。

【0007】

本発明の請求項2による音声モデルの雑音適応化システムは、請求項1において、前記クラスタリング手段は、雑音対信号比条件に合わせて前記雑音を音声に重畳することにより前記雑音重畳音声を作成し、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い、各雑音重畳音声のガウス分布モデルを作成し、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより、クラスタリング結果を得ることを特徴とする。こうすることにより、雑音重畳音声に対して、クラスタリング処理を行うことができる。

【0008】

本発明の請求項 3 による音声モデルの雑音適応化システムは、請求項 1 又は 2 において、前記選択手段は、前記パラメータ抽出手段によって抽出された音声特徴パラメータに対する尤度が最大となるモデルを選択することを特徴とする。尤度が最大となるモデルを選択することにより、音声認識精度を高めることができる。

【0009】

本発明の請求項 4 による音声モデルの雑音適応化システムは、請求項 3 において、前記選択手段は、前記木構造の雑音音声モデル空間について、上層から下層に向かってたどることによってモデルを選択することを特徴とする。木構造の上層から下層に向かってたどることによって最適なモデルを選択することができる。

【0010】

本発明の請求項 5 による音声モデルの雑音適応化システムは、請求項 1 乃至 4 のいずれか 1 項において、前記線形変換手段は、前記選択手段によって選択されたモデルに基づいて尤度が大きくなるように線形変換を行うことを特徴とする。線形変換を行うことにより、尤度を最大にすることができる。

【0011】

本発明の請求項 6 による音声モデルの雑音適応化方法は、クリーンな音声データを用いて学習した不特定雑音用の音声モデルを、認識対象となる雑音環境の音声に適応化する音声モデルの雑音適応化方法であって、雑音が重畳された雑音重畳音声のクラスタリングを行うクラスタリングステップと、このクラスタリングステップによるクラスタリング結果に基づいて木構造の雑音音声モデル空間を作成する音声モデル空間作成ステップと、音声認識対象となる入力雑音音声の音声特徴パラメータを抽出するパラメータ抽出ステップと、前記音声モデル空間作成ステップにおいて作成された木構造の雑音音声モデル空間から最適なモデルを選択する選択ステップと、この選択ステップにおいて選択されたモデルについて尤度が更に大きくなるように線形変換を行う線形変換ステップとを含むことを特徴とする。このように、クラスタリング過程とモデル学習過程とを、一貫して雑音重畳音声に対して行うことによって、入力音声に対する音声モデル系列の推定を

より正確に行うことができる。

【0012】

本発明の請求項7による音声モデルの雑音適応化方法は、請求項6において、前記クラスタリングステップにおいては、雑音対信号比条件に合わせて前記雑音を音声に重畳することにより前記雑音重畳音声を作成し、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い、各雑音重畳音声のガウス分布モデルを作成し、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより、クラスタリング結果を得ることを特徴とする。こうすることにより、雑音重畳音声に対して、クラスタリング処理を行うことができる。

【0013】

本発明の請求項8による音声モデルの雑音適応化方法は、請求項6又は7において、前記選択ステップにおいては、前記パラメータ抽出ステップにおいて抽出された音声特徴パラメータに対する尤度が最大となるモデルを選択することを特徴とする。尤度が最大となるモデルを選択することにより、音声認識精度を高めることができる。

【0014】

本発明の請求項9による音声モデルの雑音適応化方法は、請求項8において、前記選択ステップにおいては、前記木構造の雑音音声モデル空間について、上層から下層に向かってたどることによってモデルを選択することを特徴とする。木構造の上層から下層に向かってたどることによって最適なモデルを選択することができる。

【0015】

本発明の請求項10による音声モデルの雑音適応化方法は、請求項6乃至9のいずれか1項において、前記線形変換ステップにおいては、前記選択ステップにおいて選択されたモデルに基づいて尤度が大きくなるように線形変換を行うことを特徴とする。線形変換を行うことにより、尤度を最大にすることができる。

【0016】

本発明の請求項11による音声認識雑音適応化プログラムは、コンピュータを制御することにより、クリーンな音声データを用いて学習した不特定雑音用の音

声モデルを、認識対象となる雑音環境の音声に適応化するための音声認識雑音適応化プログラムであって、雑音が重畳された雑音重畳音声のクラスタリングを行うクラスタリングステップと、このクラスタリングステップによるクラスタリング結果に基づいて木構造の雑音音声モデル空間を作成する音声モデル空間作成ステップと、音声認識対象となる入力雑音音声の音声特徴パラメータを抽出するパラメータ抽出ステップと、前記音声モデル空間作成ステップにおいて作成された木構造の雑音音声モデル空間から最適なモデルを選択する選択ステップと、この選択ステップにおいて選択されたモデルについて尤度が更に大きくなるように線形変換を行う線形変換ステップとを含むことを特徴とする。このように、クラスタリング過程とモデル学習過程とを、一貫して雑音重畳音声に対して行うことによって、入力音声に対する音声モデル系列の推定をより正確に行うことができる。

【0017】

本発明の請求項12による音声認識雑音適応化プログラムは、請求項11において、前記クラスタリングステップにおいては、雑音対信号比条件に合わせて前記雑音を音声に重畳することにより前記雑音重畳音声を作成し、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い、各雑音重畳音声のガウス分布モデルを作成し、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより、クラスタリング結果を得ることを特徴とする。こうすることにより、雑音重畳音声に対して、クラスタリング処理を行うことができる。

【0018】

本発明の請求項13による音声認識雑音適応化プログラムは、請求項11又は12において、前記選択ステップにおいては、前記パラメータ抽出ステップにおいて抽出された音声特徴パラメータに対する尤度が最大となるモデルを選択することを特徴とする。尤度が最大となるモデルを選択することにより、音声認識精度を高めることができる。

【0019】

本発明の請求項14による音声認識雑音適応化プログラムは、請求項13において、前記選択ステップにおいては、前記木構造の雑音音声モデル空間について

、上層から下層に向かってたどることによってモデルを選択することを特徴とする。木構造の上層から下層に向かってたどることによって最適なモデルを選択することができる。

【0020】

本発明の請求項15による音声認識雑音適応化プログラムは、請求項11乃至14のいずれか1項において、前記線形変換ステップにおいては、前記選択ステップにおいて選択されたモデルに基づいて尤度が大きくなるように線形変換を行うことを特徴とする。線形変換を行うことにより、尤度を最大にすることができる。

【0021】

要するに本発明では、雑音データベースを用いて、SNR条件に合わせて、クリーンな音声に雑音を重畳する（後述するステップS1）。この雑音重畳音声に対しクラスタリングを行って木構造雑音重畳音声モデル空間を構築する。雑音重畳音声モデル空間において、各木構造ノードに属する雑音をクリーンな音声に重畳し、雑音重畳音声のモデルを構築する（後述するステップS3）。雑音重畳音声の木構造モデル空間において、尤度を計算し（後述するステップS4）、上から下にたどり最適なモデルを選択する（後述するステップS7）。このように選択された適応化音声モデル系列のモデルパラメータに基づいて尤度が最大となるように線形変換を行う（後述するステップS8）。

【0022】

つまり、モデル学習過程のみならず、クラスタリング過程においても雑音重畳音声に対して処理を行う。このように、クラスタリング過程とモデル学習過程とを、一貫して雑音重畳音声に対して行うことによって、最尤の雑音重畳音声モデルを学習できる。したがって、高い認識精度を達成できる。

【0023】

【発明の実施の形態】

次に、図面を参照して本発明の実施の形態について説明する。なお、以下の説明において参照する各図においては、他の図と同等部分に同一符号が付されている。

本発明においては、雑音音声モデル空間を、雑音対信号比すなわち SNR (signal-to-noise ratio) と音質とによって、木構造的に作成しておく。木構造で雑音特性を表すことにより、木構造の上層では雑音特性の大局的な特徴、下層では局所的な特徴を表現するモデルが得られる。この木構造を上から下にたどりルートからトップダウン方式で最適なモデルを選択することにより、最適な雑音区分空間を選択できる。

【0024】

そして、クラスタリング過程とモデル学習過程とを、一貫して雑音重畳音声に対して行うことによって、最尤の雑音重畳音声モデルを学習でき、認識精度を高めることができる。

(本システムの構成)

この処理を実現するための構成について図1を参照して説明する。同図は、本発明による雑音適応化システムの実施の一形態を示すブロック図である。同図に示されているように、本実施形態による雑音適応化システムは、木構造モデル記憶部1と、特徴抽出部2と、音声認識部3と、モデル選択判定部4と、モデル線形変換適応部5と、認識結果保存部6とを含んで構成されている。

【0025】

木構造モデル記憶部1は、雑音重畳音声のクラスタリングの結果によって、雑音重畳音声HMMを二段階 (SNR 及び雑音性質) の木構造に作成し、保存する部分である。

特徴抽出部2は、入力された音声データを分析して特徴ベクトルに変換する分析部分である。

【0026】

音声認識部3は、特徴ベクトルの時系列に変換された上記入力音声データを $Viterbi$ アルゴリズムによって、尤度関数が最大となるモデル系列を求める部分である。

モデル選択判定部4は、木構造モデル記憶部1に記憶されているモデルの中から、尤度が最大となるように最適なモデルを選択する部分である。

【0027】

モデル線形変換適応部 5 は、モデル選択判定部 4 によって選択されるモデルで、尤度が最大化するように線形変換を行う部分である。

認識結果保存部 6 は、音声の認識結果を保存する部分である。

(本システムの動作)

以上の構成からなる本システムの動作について、図 2 及び図 3 を参照して説明する。図 2 は、本システムの動作の流れに従って図 1 中の各部 1 ～ 6 を並べ替えた機能ブロック図である。図 3 は本システムによる処理手順を示すフロー図である。

【0028】

本システムを使った音声認識を実施する手順は、以下のステップ S 1 ～ S 9 のようになる。

ステップ S 1 (雑音が重畳された音声の作成過程)：雑音データベースを用いて、SNR 条件に合わせて、クリーンな音声に重畳し、雑音が重畳された音声すなわち雑音重畳音声を作成する。なお、本例では、 $SNR = 5, 10, 15 \text{ dB}$ とする。

【0029】

ステップ S 2 (雑音重畳音声の平均差引き適応過程)：ステップ S 1 で作成される雑音重畳音声に対し、CMS (Cepstral Mean Subtraction) を適用する。CMS は、音声ケプストラム (Cepstral) の平均を差引く手法である。すなわち、一定区間の音声データに対し、あらゆるフレームのケプストラムの平均値を計算し、フレームごとにこの平均値を差引く処理が行われる。ケプストラムとは、フーリエ変換によって求められたパワースペクトルの対数値をさらにフーリエ変換したものである。なお、この CMS については、文献「Furui: Cepstral Analysis Technique For Automatic Speaker Verification, IEEE Transaction on Acoustical Speech and Signal Processing, Vol. ASSP-29, pp. 254-272, 1981 年」に記載されている。

【0030】

ステップS3（雑音重畳音声モデルの作成過程）：各雑音重畳音声の混合ガウス分布モデル（GMM）をBaum-Welchアルゴリズムによって作成する。Baum-Welchアルゴリズムでは、適当な初期値から始めて、繰り返しによって最適値に近づくというアプローチをとる。なお、このBaum-Welchアルゴリズムは、文献「中川聖一著：“確率モデルによる音声認識”、電子情報通信学会、1988」に記載されている。

【0031】

ステップS4（雑音重畳音声のクラスタリング過程）：次に、ガウス分布モデルを用いて、各雑音重畳音声間の尤度を計算し、尤度行列を作成する。この尤度行列に基づいてSPLIT法を用いて、逐次的に雑音重畳音声のクラスタリングを行う。このSPLIT方法ではひずみが最大となるクラスタを順次分割する。よって任意の数のクラスタを作成できる。クラスタ数さえ与えれば、完全に自動的にクラスタリングの結果が得られる。なお、このSPLIT法は、文献「菅村 他、音声研究会資料、S82-64、1982年」に記載されている。

【0032】

ステップS5（区分線形変換適応への適用）：上記のステップS4により、雑音重畳音声の木構造クラスタリング結果が得られる。このクラスタリング結果を木構造モデル記憶部1に保存する。このクラスタリング結果は、木構造で特性を表しているので、木構造の上層では雑音重畳音声の大局的な特徴を、下層では局所的な特徴を、それぞれ表現している。

【0033】

このクラスタリング結果を、区分線形変換手法に適用する。この区分線形変換手法は、文献「張他、日本音響学会2002年秋季発表会、pp. 29-30」に記載されている。具体的には以下のステップS6～S9の処理が行われる。

ステップS6（特徴量抽出過程）：特徴抽出部2において、認識対象雑音音声データの特徴量を抽出する。特徴量抽出は、入力された音声データを一定フレーム毎に線形予測符号化（Linear Prediction Coding；LPC）分析し、ケプストラム又は Δ ケプストラムなどの特徴パラメータベクトルの時系列を特徴パラメータ列として得る。

【0034】

ステップS7（最適モデル選択）：最適モデル選択について、図4を参照して説明する。図4において、一番上のノード（root）はクリーン音声モデルを表す。その下にはSNRごとにN個のモデルがある。N個のモデルとは、モデルSNR-1～モデルSNR-Nである。これらN個のモデルSNR-1～モデルSNR-Nは特定SNR条件での全雑音を混合した音声から学習したモデルを表す。

【0035】

そのさらに下にある子ノードはクラスタリング結果によって、選択された一部の雑音種類を混合した音声データから学習したモデルを表す。この木構造の一番下にあるのはある特定雑音（1種類だけ）を混合した音声から学習したモデルである。木構造の上層では雑音特性の大局的な特徴、下層では局所的な特徴を表現する。

【0036】

認識するときは、まずrootのクリーン不特定モデルを用いて、ステップS4で得られた特徴パラメータ列によって、尤度を計算する。これは図1中の音声認識部3によって行われる。

次に、rootの下各モデルを用いて、同じく音声認識部3によって尤度を計算する。以上で得られた尤度の値を用いて、モデル選択判定部4によって最適モデルを選択する。具体的に以下のように処理する。rootのクリーン不特定モデルの尤度より高いモデルを保留する。これらのSNR条件でさらに下へ各子ノードのモデルを用いて尤度を計算する。2つの子ノードモデルと親ノードの尤度を比較し、尤度最大のモデルは子ノードモデルであれば、さらに下へたどる。親ノードの方が一番高ければ、計算を停止し、この親ノードを最適なノードだと判断する。

【0037】

図4には、探索パスが実線で表されている。この計算を繰り返すと最適な空間を見つけることができる。また、各SNR条件の尤度最大となるモデルの尤度を比較し、最大となるモデルが全体の雑音音声空間での最適モデルだとする。例え

ば、図4中のSNR-1条件では4番ノードが尤度最大である。また、同図中のSNR-N条件では5番ノードが尤度最大である。そして、これら各SNR条件において最大のモデルの中からさらに尤度同士を比較し、選択する。

【0038】

ステップS8（線形変換）：選択されるモデルについて、モデル線形変換適応部5によってさらに尤度が最大化するように線形変換（Maximum Likelihood Linear Regression；以下MLLRと略称する）を行う。この線形変換については、文献「Mean and variance adaptation within the MLLR framework」（M. J. F. Gales et al.、Computer Speech and Language、pp. 249-264、1996年）に記載されている。具体的には認識結果の音素列を用いて、線形変換の行列を尤度最大基準に基づいて推定し、HMMのガウス分布の平均値と分散値とを線形変換により適応化する。

【0039】

ステップS9（再認識）：音声認識結果を出力する場合は、ステップS8で得られたモデルを用いて、音声認識部3によって再認識し、認識結果を認識結果保存部6に保存する。

（本システムの効果）

以上のように本発明では、雑音データベースを用いて、雑音を音声に重畳させる、雑音重畳音声のモデルを学習する。各雑音モデル間の距離を計算し、雑音のクラスタリングを行う。雑音のクラスタリングの結果に基づいて、木構造的に音声モデルを作成する。

【0040】

まず雑音重畳音声についてSNRでクラスタ化し、次にSNR条件ごとに木構造モデルが用意でき、木構造雑音音声モデル空間を作成しておく。特徴抽出過程で、認識対象となる入力雑音音声进行分析して特徴パラメータ列を抽出し、木構造雑音音声モデル空間から最適なモデルを選択する。選ばれた雑音音声モデル空間から尤度がさらに最大化するように線形変換を行う。

【0 0 4 1】

以上述べたように、本発明の雑音適応化システムにおいては、認識対象となる雑音音声を用いて、不特定話者用音声モデルを適応化している。雑音重畳音声の木構造モデル空間を作成し、上から下にたどり最適なモデルを選択し、さらに尤度がさらに最大化するように線形変換を行って、適応化モデルを作成する。

ここで、発明者は、雑音に乗っている対話音声について、本システムによる音声認識の効果を調べた。以下、この実験例を述べる。

【0 0 4 2】

実験で使用した音声HMMは、`tree-based clustering`により状態共有化を行った不特定話者文脈依存音素HMMである。特徴量としては、MFCC (`Mel Frequency Cepstral Coefficients`) 12次元、その微分12次元、対数パワーの1次微分の計25次元を利用した。「`Mel Frequency`」とは、人間が音を聴く感度に合わせて作られた値である。その音がどれくらい人間の耳によく聴こえるのかを表すときに良く使用される。このMFCCは以下の手順で生成される。すなわち、音声波形のデータに対し、離散フーリエ変換を行い、その値を対数表現に変換する。そして、それを逆離散フーリエ変換し、そこでできた波形を一定間隔ごとに取出したものがMFCCである。

【0 0 4 3】

本システムの効果について図5を参照して説明する。同図には、不特定音声HMMを用いた単語正解精度（ベースライン）と、この発明の実施例の方法により適応化された音声HMMを用いた単語正解精度（本手法）とが示されている。同図中の縦軸は単語正解精度（％）である。同図中の横軸はSNR（dB）である。また、同図中の散点模様がベースライン、縦縞模様が本システムによる手法である。

【0 0 4 4】

同図に示されている結果から、本発明による方法が従来方法に比べて有効であることがわかる。この場合、ベースラインに比べ、単語誤り率は36.1％低下した。

(音声モデルの雑音適応化方法)

上述した雑音適応化システムにおいては、以下のような雑音適応化方法が実現されている。すなわち、クリーンな音声データを用いて学習した不特定雑音用の音声モデルを、認識対象となる雑音環境の音声に適応化する音声モデルの雑音適応化方法であり、雑音が重畳された雑音重畳音声のクラスタリングを行うクラスタリングステップと、このクラスタリングステップによるクラスタリング結果に基づいて木構造の雑音音声モデル空間を作成する音声モデル空間作成ステップと、音声認識対象となる入力雑音音声の音声特徴パラメータを抽出するパラメータ抽出ステップと、上記音声モデル空間作成ステップにおいて作成された木構造の雑音音声モデル空間から最適なモデルを選択する選択ステップと、この選択ステップにおいて選択されたモデルについて尤度が更に大きくなるように線形変換を行う線形変換ステップとを含む雑音適応化方法が実現されている。このように、クラスタリング過程とモデル学習過程とを、一貫して雑音重畳音声に対して行うことによって、入力音声に対する音声モデル系列の推定をより正確に行うことができる。

【0045】

また、上記クラスタリングステップにおいては、雑音対信号比条件に合わせて上記雑音を音声に重畳することにより上記雑音重畳音声を作成し、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い、各雑音重畳音声のガウス分布モデルを作成し、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより、クラスタリング結果を得る。こうすることにより、雑音重畳音声に対して、クラスタリング処理を行うことができる。

【0046】

さらに、上記選択ステップにおいては、上記パラメータ抽出ステップにおいて抽出された音声特徴パラメータに対する尤度が最大となるモデルを選択する。尤度が最大となるモデルを選択することにより、音声認識精度を高めることができる。

そして、上記選択ステップにおいては、上記木構造の雑音音声モデル空間について、上層から下層に向かってたどることによってモデルを選択する。木構造の

上層から下層に向かってたどることによって最適なモデルを選択することができる。

【0047】

なお、上記線形変換ステップにおいては、上記選択ステップにおいて選択されたモデルに基づいて尤度が大きくなるように線形変換を行うことにより、尤度を最大にすることができる。

(音声モデルの雑音適応化プログラム)

ところで、図3に示されている処理を実行するためのプログラムを用意し、これを用いてコンピュータを制御することにより、上記と同様の効果が得られる。このプログラムは、コンピュータを制御することにより、クリーンな音声データを用いて学習した不特定雑音用の音声モデルを、認識対象となる雑音環境の音声に適応化するための音声認識雑音適応化プログラムであり、雑音が重畳された雑音重畳音声のクラスタリングを行うクラスタリングステップと、このクラスタリングステップによるクラスタリング結果に基づいて木構造の雑音音声モデル空間を作成する音声モデル空間作成ステップと、音声認識対象となる入力雑音音声の音声特徴パラメータを抽出するパラメータ抽出ステップと、上記音声モデル空間作成ステップにおいて作成された木構造の雑音音声モデル空間から最適なモデルを選択する選択ステップと、この選択ステップにおいて選択されたモデルについて尤度が更に大きくなるように線形変換を行う線形変換ステップとを含んでいる。このように、クラスタリング過程とモデル学習過程とを、一貫して雑音重畳音声に対して行うことによって、入力音声に対する音声モデル系列の推定をより正確に行うことができる。

【0048】

また、上記クラスタリングステップにおいては、雑音対信号比条件に合わせて上記雑音を音声に重畳することにより上記雑音重畳音声を作成し、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い、各雑音重畳音声のガウス分布モデルを作成し、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより、クラスタリング結果を得る。こうすることにより、雑音重畳音声に対して、クラスタリング処理を行うことができる。

【0049】

さらに、上記選択ステップにおいては、上記パラメータ抽出ステップにおいて抽出された音声特徴パラメータに対する尤度が最大となるモデルを選択する。尤度が最大となるモデルを選択することにより、音声認識精度を高めることができる。

そして、上記選択ステップにおいては、上記木構造の雑音音声モデル空間について、上層から下層に向かってたどることによってモデルを選択する。木構造の上層から下層に向かってたどることによって最適なモデルを選択することができる。

【0050】

なお、上記線形変換ステップにおいては、上記選択ステップにおいて選択されたモデルに基づいて尤度が大きくなるように線形変換を行うことにより、尤度を最大にすることができる。

以上のプログラムを記録するための記録媒体には、図示されていない半導体メモリ、磁気ディスク、光ディスク等の他、種々の記録媒体を用いることができる。

【0051】**【発明の効果】**

以上説明したように、本発明の請求項1、6、11によれば、クラスタリング過程とモデル学習過程とを、一貫して雑音重畳音声に対して行うことによって、多くの雑音データについて最適にクラスタリング処理をすることができると共に、入力音声に対する音声モデル系列の推定をより正確に行うことができるという効果がある。

【0052】

また、本発明の請求項2、7、12によれば、雑音対信号比条件に合わせて雑音を音声に重畳し、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い、各雑音重畳音声のガウス分布モデルを作成し、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより、雑音重畳音声に対して、クラスタリング処理を行うことができるという効果がある。

【0053】

本発明の請求項3、8、13によれば、抽出された音声特徴パラメータに対する尤度が最大となるモデルを選択することにより、音声認識精度を高めることができるという効果がある。

本発明の請求項4、9、14によれば、木構造の雑音音声モデル空間について、上層から下層に向かってたどることによってモデルを選択することにより、最適なモデルを選択することができるという効果がある。

【0054】

本発明の請求項5、10、15によれば、選択されたモデルに基づいて尤度が大きくなるように線形変換を行うことにより、尤度を最大にすることができるという効果がある。

【図面の簡単な説明】**【図1】**

本発明の実施の一形態による音声モデルの雑音適応化システムの構成を示すブロック図である。

【図2】

本システムの動作の流れに従って図1中の各部を並べ替えた機能ブロック図である。

【図3】

本システムによる処理手順を示すフロー図である。

【図4】

木構造雑音音声モデル空間における最適モデルの選択処理を示す概念図である。

【図5】

本システムにより適応化された音声HMMを用いた単語正解精度を示す図である。

【符号の説明】

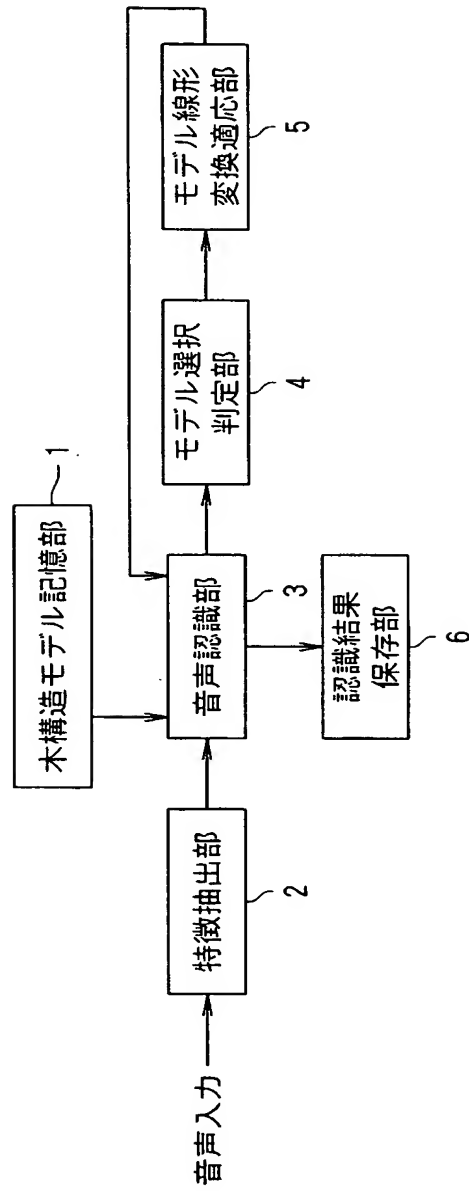
- 1 木構造モデル記憶部
- 2 特徴抽出部

- 3 音声認識部
- 4 モデル選択判定部
- 5 モデル線形変換適応部
- 6 認識結果保存部

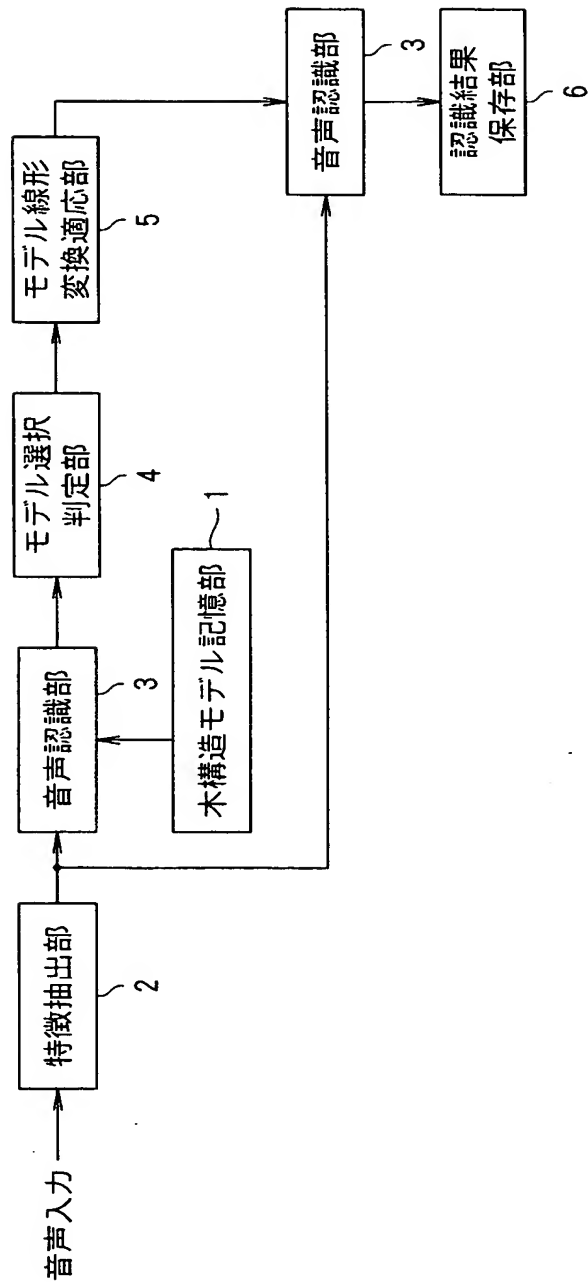
【書類名】

図面

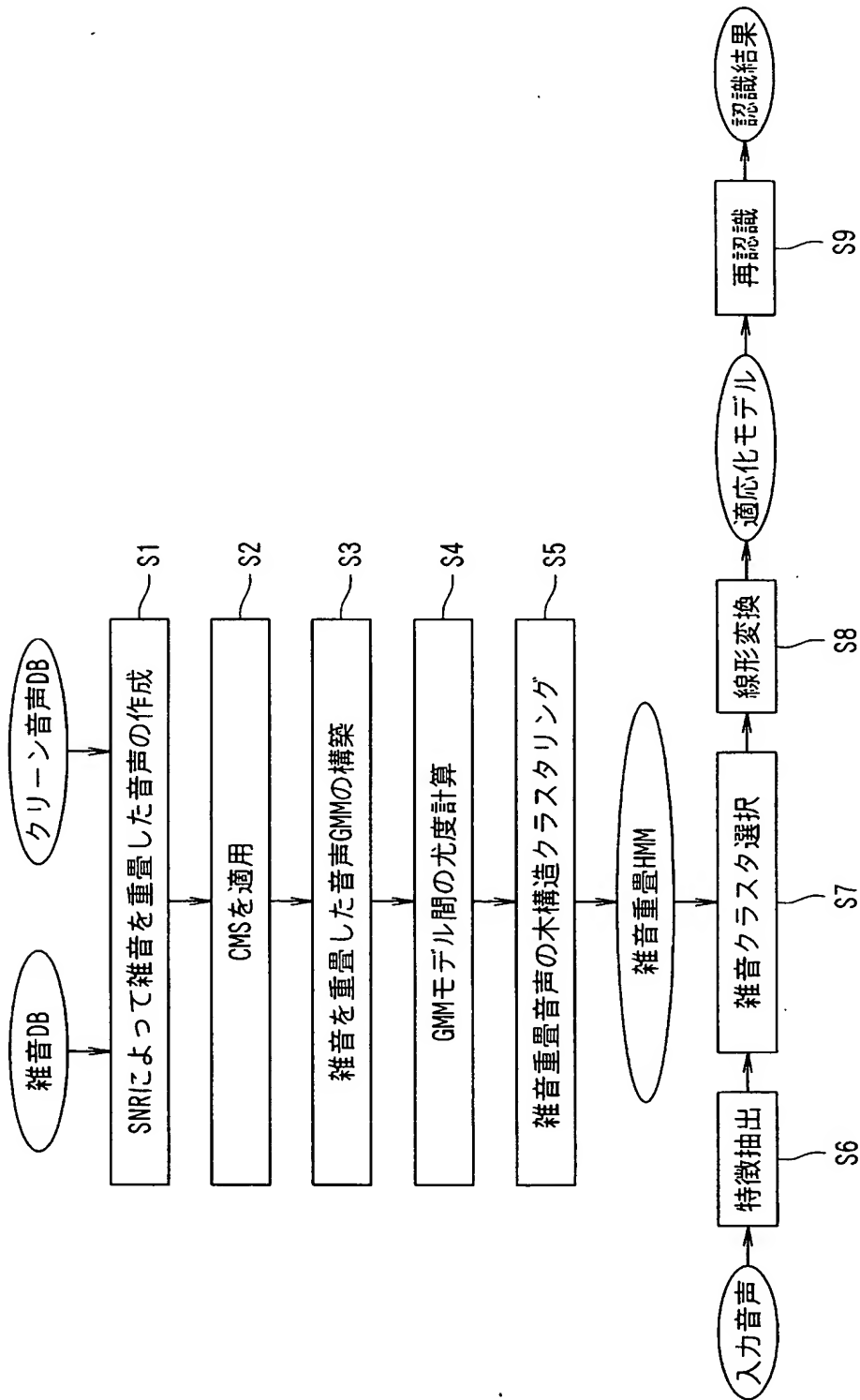
【図 1】



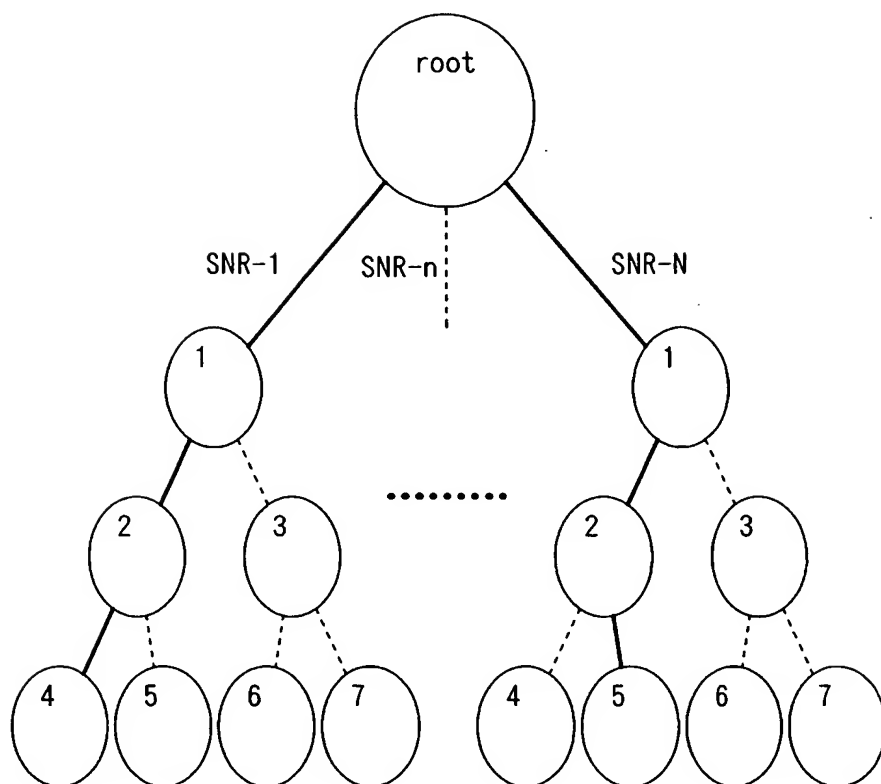
【図 2】



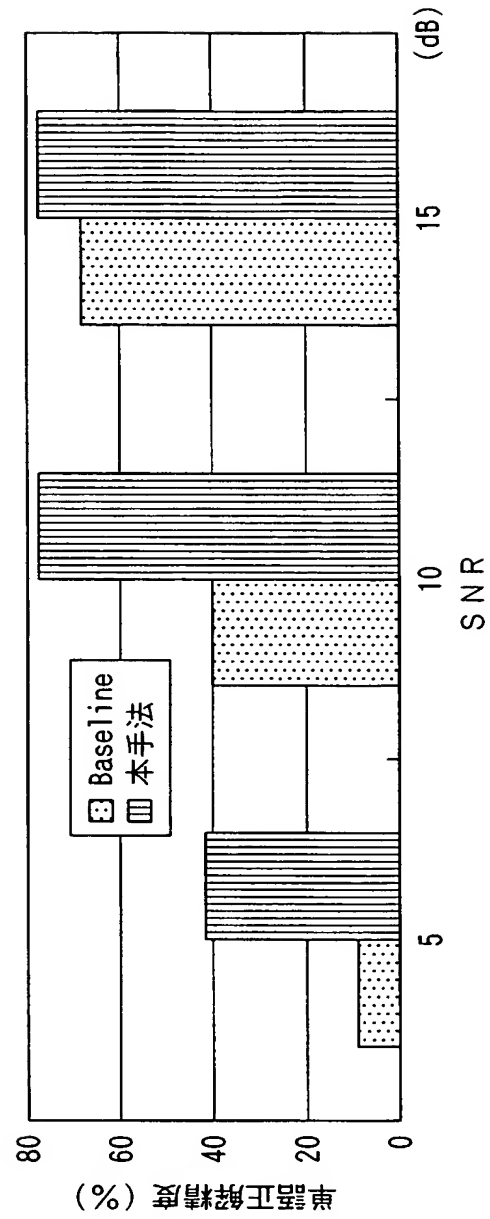
【図 3】



【図 4】



【図 5】



【書類名】 要約書

【要約】

【課題】 多くの雑音データについて最適にクラスタリング処理をすることができると共に、入力音声に対する音声モデル系列の推定をより正確に行う。

【解決手段】 雑音対信号比条件に合わせて雑音を音声に重畳することにより雑音重畳音声を作成し（ステップ S 1）、作成された雑音重畳音声について音声ケプストラムの平均値を差引く処理を行い（ステップ S 2）、各雑音重畳音声のガウス分布モデルを作成し（ステップ S 3）、各雑音重畳音声間の尤度を計算して尤度行列を作成することにより（ステップ S 4）、クラスタリング結果を得る。最適なモデルを選択し（ステップ S 7）、尤度が最大となるように線形変換する（ステップ S 8）。

【効果】 雑音重畳音声に対して、クラスタリング過程とモデル学習過程とを行うので、多くの雑音データについてのクラスタリング処理、正確な音声モデル系列推定を実現できる。

【選択図】 図 3

特願 2 0 0 3 - 0 6 6 9 3 3

出 願 人 履 歴 情 報

識別番号 [3 9 2 0 2 6 6 9 3]

1. 変更年月日 2 0 0 0 年 5 月 1 9 日

[変更理由] 名称変更

住所変更

住 所 東京都千代田区永田町二丁目 1 1 番 1 号
氏 名 株式会社 エヌ・ティ・ティ・ドコモ

特願 2 0 0 3 - 0 6 6 9 3 3

出 願 人 履 歴 情 報

識別番号

[5 0 2 3 4 7 7 8 8]

1. 変更年月日

2 0 0 2 年 9 月 2 5 日

[変更理由]

新規登録

住 所

東京都世田谷区新町二丁目 3 6 番 6 号

氏 名

古井 貞熙